

# Profiling Protein Kinases and Other ATP Binding Proteins in Arabidopsis Using Acyl-ATP Probes<sup>\*</sup>

Joji Grace Villamor<sup>‡</sup>, Farnusch Kaschani<sup>‡§</sup>, Tom Colby<sup>¶</sup>, Julian Oeljeklaus<sup>§</sup>, David Zhao<sup>||</sup>, Markus Kaiser<sup>§</sup>, Matthew P. Patricelli<sup>||</sup>, and Renier A. L. van der Hoorn<sup>‡\*\*</sup>

Many protein activities are driven by ATP binding and hydrolysis. Here, we explore the ATP binding proteome of the model plant *Arabidopsis thaliana* using acyl-ATP (AcATP)<sup>1</sup> probes. These probes target ATP binding sites and covalently label lysine residues in the ATP binding pocket. Gel-based profiling using biotinylated AcATP showed that labeling is dependent on pH and divalent ions and can be competed by nucleotides. The vast majority of these AcATP-labeled proteins are known ATP binding proteins. Our search for labeled peptides upon in-gel digest led to the discovery that the biotin moiety of the labeled peptides is oxidized. The in-gel analysis displayed kinase domains of two receptor-like kinases (RLKs) at a lower than expected molecular weight, indicating that these RLKs lost the extracellular domain, possibly as a result of receptor shedding. Analysis of modified peptides using a gel-free platform identified 242 different labeling sites for AcATP in the Arabidopsis proteome. Examination of each individual labeling site revealed a preference of labeling in ATP binding pockets for a broad diversity of ATP binding proteins. Of these, 24 labeled peptides were from a diverse range of protein kinases, including RLKs, mitogen-activated protein kinases, and calcium-dependent kinases. A significant portion of the labeling sites could not be assigned to known nucleotide binding sites. However, the fact that labeling could be competed with

ATP indicates that these labeling sites might represent previously uncharacterized nucleotide binding sites. A plot of spectral counts against expression levels illustrates the high specificity of AcATP probes for protein kinases and known ATP binding proteins. This work introduces profiling of ATP binding activities of a large diversity of proteins in plant proteomes. The data have been deposited in ProteomeXchange with the identifier PXD000188. *Molecular & Cellular Proteomics* 12: 10.1074/mcp.M112.026278, 2481–2496, 2013.

ATP binding and hydrolysis are the driving processes in all living organisms. Hundreds of cellular proteins are able to bind and hydrolyze ATP to unfold proteins, transport molecules over membranes, or phosphorylate small molecules or proteins. Proteins with very different structures are able to bind ATP. A large and important class of ATP binding proteins is that of the kinases, which transfer the gamma phosphate from ATP to substrates. Kinases, and particularly protein kinases, play pivotal roles in signaling and protein regulation.

The genome of the model plant *Arabidopsis thaliana* encodes for over 1099 protein kinases and hundreds of other ATP binding proteins (1, 2). Protein kinases are involved in nearly all signaling cascades and regulate processes ranging from cell cycle to flowering and from immunity to germination. Many protein kinases in plants are receptor-like kinases (RLKs), often carrying extracellular leucine-rich repeats (LRRs). The RLK class contains at least 610 members (3), including famous examples such as receptors involved in development (e.g. BRI1, ER, CLV1) and immunity (e.g. FLS2, EFR). Other important classes are mitogen-activated protein (MAP) kinases (MPKs) (20 different members), MPK kinase kinases (MAP3Ks) (60 different members (4)), and calcium-dependent protein kinases (CPKs) (34 different members (5)). Because of their diverse and important roles, protein kinases have been intensively studied in plant science. The current approach is to study protein kinases individually—a daunting task, considering the remaining hundreds of uncharacterized protein kinases. New approaches are necessary in order to study protein kinases and other ATP binding proteins globally rather than individually.

From the <sup>‡</sup>Plant Chemetics Laboratory, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany; <sup>§</sup>Zentrum für Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Essen, Germany; <sup>¶</sup>Proteomics Service Center, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany; <sup>||</sup>ActivX Biosciences, 11025 North Torrey Pines Road, La Jolla, California 92037

Received December 4, 2012, and in revised form, May 9, 2013

Published, MCP Papers in Press, May 29, 2013, DOI 10.1074/mcp.M112.026278

<sup>1</sup> The abbreviations used are: AcATP, acyl-ATP; BHAc, biotin acyl; BCCP, biotin carboxyl carrier protein; BHAcATP, biotin-acyl-ATP; CPK, calcium-dependent protein kinase; DBAcATP, desthiobiotin-acyl-ATP; HSP, heat shock protein; LRR, leucine-rich repeat; MAP, mitogen-activated protein; MAP3K, mitogen-activated protein kinase kinase; MCCA, 3-methylcrotonyl-CoA carboxylase; MPK, mitogen-activated protein kinase; MW, molecular weight; OBHAc, oxidized BHAc; PERK, proline-rich extension-like receptor kinase; PGK, phosphoglycerate kinase; RLK, receptor-like kinase; VIK, VH1-interacting kinase.

ATP binding activities of protein kinases and other proteins can be detected globally by acyl-ATP (AcATP) probes (6, 7) (Fig. 1A). AcATP binds to the ATP pocket of ATP binding proteins and places the acyl group in close proximity to conserved lysine residues in the ATP binding pocket. The acyl phosphonate moiety serves as an electrophilic warhead that can be nucleophilically attacked by the amino group of the lysine, resulting in a covalent attachment of the acyl reporter of the AcATP probe on the lysine and a concomitant release of ATP. The reporter tag is usually a biotin to capture and identify the labeled proteins. Labeled proteins can be displayed on protein blots using streptavidin-HRP. However, because AcATP labels many ATP binding proteins and protein kinases are of relatively low abundance, mass spectrometry is more often used to identify and quantify labeling with AcATP probes. The analysis is preferably done using Xsite, a procedure that involves trypsinization of the entire labeled proteome, followed by analysis of the biotinylated peptides rather than the biotinylated proteins (8). This “KiNativ” approach provides enough depth and resolving power to monitor ~160 protein kinases in a crude mammalian proteome (7). Of the 518 human protein kinases (9), 394 (76%) have been detected via AcATP labeling (6).

KiNativ has mostly been used to validate targets of human drugs that target protein kinases using competitive labeling experiments. This approach has been used to identify selective inhibitors of, for example, Parkinson’s disease protein kinase LRRK2 (10), the BMK1 and JNK MAP kinases (11, 12), and the mTOR kinase (13). Importantly, the correlation of the biological activity of protein-kinase-inhibiting drugs with inhibitor affinity detected using KiNativ is better than that achieved when affinities are determined by assays using heterologously expressed protein kinases (7). This improved correlation illustrates that assays in the native environment provide a more realistic measure of protein kinase function.

In addition to characterizing inhibitors selectively, AcATP probes can also display differential ATP binding activities of protein kinases. For example, labeling with AcATP probes during infection with dengue virus displayed a 2- to 8-fold activation of a DNA-dependent protein kinase (14). Similarly, AcATP labeling revealed an unexpected Raf kinase activation in extracts upon protein kinase inhibitor treatment (7). In conclusion, profiling with AcATP probes is a powerful approach for monitoring protein kinases and offers unprecedented opportunities to identify selective protein kinase inhibitors and discover protein kinases with differential ATP binding activities.

In this work, we introduce AcATP profiling of plant proteomes. In addition to the analysis of labeled peptides, we characterized labeling using gel-based approaches and discovered that biotin is often oxidized in this procedure. We also performed an in-depth analysis of labeling sites in proteins other than protein kinases, which had not been done before. We discuss labeling outside known nucleotide binding pock-

ets and investigate the correlation of labeling sites with protein abundance. We describe 63 labeling sites of known nucleotide binding pockets, of which 24 represent a remarkable diversity of protein kinases, including several LRR-RLKs. This work launches a new approach to study ATP binding proteins in plant science.

## EXPERIMENTAL PROCEDURES

**Plant Growth Conditions**—*Arabidopsis thaliana* ecotype Col-0 was grown in a climate cabinet in a 10-h light regime at 22 °C and 60% relative humidity.

**Probes**—Biotin-AcATP was synthesized following the procedures described by Patricelli *et al.* (6). Briefly, *N*-(+)-biotinyl-6-amino-hexanoic acid (30 mg, 0.085 mmol) was suspended under an argon atmosphere in anhydrous dioxane/*N,N*-dimethylformamide/dimethyl sulfoxide (1:1:1, 3 ml) and cooled to 0 °C. To this ice-cold solution were then added triethylamine (47  $\mu$ l, 0.34 mmol) and *iso*-butyl chloroformate (33  $\mu$ l, 0.255 mmol), resulting in the appearance of a precipitate. The resulting suspension was stirred at 0 °C for 20 min, allowed to warm up to room temperature, and then stirred for an additional 1.5 h. ATP triethylammonium salt (69 mg, 0.085 mmol) dissolved in anhydrous dimethyl sulfoxide (1 ml) was added to the reaction mixture. After 18 h, the reaction was stopped by the addition of water (4 ml), and the mixture was extracted with ethyl acetate (3  $\times$  4 ml). The aqueous layer was frozen and subsequently lyophilized. The resulting solid was suspended in water (1 ml), transferred to a reverse-phase C<sub>18</sub> flash-column (LiChroprep® RP-18, 40–63  $\mu$ m, Merck) pre-equilibrated with 5% acetonitrile in water and eluted with 30% acetonitrile in water. Product-containing fractions were pooled, frozen, lyophilized, and subsequently stored at –20 °C (yield: 4.3 mg, 5.1  $\mu$ mol, 6%). The identity of the product was confirmed via liquid chromatography electrospray ionization mass spectrometry analysis on a Thermo Scientific LCQ Fleet™ electrospray ionization spectrometer equipped with an Eclipse XDB-C18 5- $\mu$ m column from Agilent, Santa Clara, CA and using a linear gradient of solvent B (5 mM NH<sub>4</sub>OAc in acetonitrile) in solvent A (5 mM NH<sub>4</sub>OAc in H<sub>2</sub>O) at a flow rate of 1 ml/min (gradient program: 0 min/10% B  $\rightarrow$  1 min/10% B  $\rightarrow$  10 min/100% B  $\rightarrow$  12 min/100% B  $\rightarrow$  15 min/10% B). At *R*<sub>t</sub> = 2.16 min, calculated for C<sub>26</sub>H<sub>40</sub>N<sub>8</sub>O<sub>16</sub>P<sub>3</sub>S<sup>–</sup> [M-H]<sup>–</sup> 845.15, we found 845.13. Desthiobiotin-AcATP is commercially available from Thermo Scientific.

**Labeling with BHAcATP**—*Arabidopsis* proteome was extracted from 4-week-old *Arabidopsis* rosette in 50 mM Tris pH 7.5. The lysate was cleared via centrifugation and subjected to gel filtration using DG10 columns (Bio-Rad). Labeling was performed in 50 mM Tris pH 7.5, 10 mM MgCl<sub>2</sub>, unless specified otherwise. The lysate was incubated with 5 or 10  $\mu$ M BHAcATP or dimethyl sulfoxide for the no-probe control at room temperature for 15 min. For inhibition experiments, the lysate was preincubated with the different nucleotide-based inhibitors at a final concentration of 10 mM for 30 min before labeling with BHAcATP for 15 min. The labeling reaction was stopped by the addition of gel loading buffer containing SDS and DTT. Proteins were separated on 12% acrylamide gel and transferred onto PVDF membranes. The protein blot was probed with streptavidin-HRP (Ultrasensitive, Sigma) and detected with enhanced chemiluminescence (SuperSignal West Femto Chemiluminescent Substrate, Thermo Scientific).

**Affinity Purification and Identification of Labeled Proteins**—After labeling of *Arabidopsis* leaf extract (1 ml 5 mg/ml protein) for 1 h, the lysate was desalted using a DG10 column (Bio-Rad). The eluate was incubated with 100  $\mu$ l neutravidin beads (Thermo Scientific) in the presence of PBS and 0.2% SDS for 1 h. The beads were washed twice with 0.2% SDS, twice with 6 M urea, and twice with water and

eluted with 50% formic acid. Proteins were separated on protein gels and stained with SYPRO Ruby (Invitrogen). Bands of interest were excised from both samples and subjected to in-gel digestion with trypsin.

**In-gel Digest and MS—LC-MS/MS** analysis was performed on an LTQ Velos (Thermo) coupled to an EasyNanoLC (Proxeon/Thermo) using a “Top 20” data-directed acquisition. LC was performed on a C18 column (10 cm × 75 μm) with a gradient of 5%–40% acetonitrile in 0.1% formic acid for 30 min, followed by a wash with 95% acetonitrile. The data-directed acquisition included active exclusion for mass (±40 ppm) and time (60 s). Spectra were searched using Mascot 2.3 using a fixed modification of cysteine (57.02 Da for carbamidomethyl) and variable modifications of methionine (15.99 Da for oxidation) and lysine (339.16 and 355.16 Da for BHAc and OBHAc labeling, respectively). Mass tolerance was set at 0.3 Da for the precursor ions and 0.4 Da for fragment ions. Up to two missed cleavages for trypsin were permitted, as the labeling would prevent cleavage at the labeled lysine. The MS2 spectra were searched against TAIR10 pep 2012 (November 2012), containing 35,386 proteins of *Arabidopsis thaliana*, supplemented with a small database with 1095 artifact proteins and a reversed decoy database of the same proteins (total: 72,962 entries). Peptides were retained with Mascot scores of >41, which is above the Mascot significance level. Hits to the decoy database within this selection were given a 0.3% false discovery rate for all peptides and a 0.7% false discovery rate for non-redundant peptides. Proteins were selected having a minimum of two different peptides, as shown in supplemental Table S1. This selection for two peptides per protein removed all the false positive hits to the decoy database. Proteins were ranked on their score and manually selected considering the occurrence of the protein in the rest of the gel and its expected molecular weight and overall score and spectral count (supplemental Table S2). Only robustly identified proteins that were highly enriched in the gel slice relative to other gel slices are reported.

**Labeling with Desthiobiotin-acyl-ATP**—Analysis of desthiobiotin-acyl-ATP (DBAcATP)-labeled Arabidopsis leaf extracts was performed as described elsewhere (6, 7). Briefly, lysates from Arabidopsis leaves were generated via bead-based agitation in kinase buffer (20 mM HEPES pH 7.8, 150 mM NaCl, 0.1% Triton X-100). The resulting lysates were gel filtered into 20 mM HEPES pH 7.8, 150 mM NaCl, 1% v/v phosphatase inhibitor mixture II (Calbiochem) to remove endogenous competing nucleotides prior to labeling with 20 μM DBAcATP probe for 15 min in the presence of 20 mM MnCl<sub>2</sub>. The labeled extracts were denatured and reduced (6 M urea, 10 mM DTT, 65 °C for 15 min) and then alkylated with iodoacetamide (40 mM, 30 min at 37 °C). Following another gel filtration step (into 10 mM ammonium bicarbonate, 2 M urea) to remove unreacted reagents and lower the urea concentration, the samples were digested with trypsin. DBAcATP-labeled peptides were enriched on streptavidin resin (Thermo) and eluted using 0.1% TFA in 50% acetonitrile. MS analysis was performed on a Thermo-LTQ linear ion trap instrument in a data-dependent mode as described for the initial characterization (7) using a mass range from 500–1800 m/z.

The MS2 spectra were searched using SEQUEST against the TAIR9 pep 20090619 database (June 2009) containing 32,769 proteins of *Arabidopsis thaliana*. MS2 searches included fixed iodoacetamide modification of cysteine (57 Da for alkylation) and variable modifications of methionine (16 Da for oxidation) and lysine (196 Da for DBAcATP labeling). Up to three missed cleavages with trypsin were allowed, and non-tryptic or half-tryptic peptides were excluded. Mass tolerance was set at 3 Da for the precursor ions and 1 Da for fragment ions. This high mass tolerance for precursor ions had to be used because most of the precursor ions are detected with multiple charges, and the mass tolerance in SEQUEST is not charge state

specific. A probability score was calculated for each labeled peptide as described previously (15). This score is based on Xcorr, delta-Cn, and peptide mass/length and compared with the distribution of false positive scores generated by searching with incorrect masses of the modification (+10 and –10 Da from the correct value). The resulting MS2 spectra were assembled in one Excel file, and all spectra with <95% probability were removed. Peptides that were detected in only one of the four MS runs were discarded. The resulting list of 242 peptides included peptides that are unique (u) in the Arabidopsis proteome or that match multiple genes (ambiguous (a)). Some peptides match multiple gene models of a single gene (isoforms (i)). These isoforms are indicated in a supplemental table but were treated as identifiers for specific genes. Supplemental Table S3 contains the lists, with 10,465 spectral counts with individual scores with probabilities of >75%. Supplemental Table S4 summarizes the spectral counts of the 242 labeled peptides with >95% probability.

**Analysis of Labeling Sites**—In order to map the labeling sites onto protein structures, we searched the Protein Data Bank database for sequences that were homologous to the labeled proteins. Sequences for which a co-crystal with a nucleotide was available were aligned with the identified proteins, and the orthologous position of the labeled lysine was indicated in the alignment. This residue was selected in the protein structure using PyMol, and the distance to the nearest phosphate of the bound nucleotide was measured.

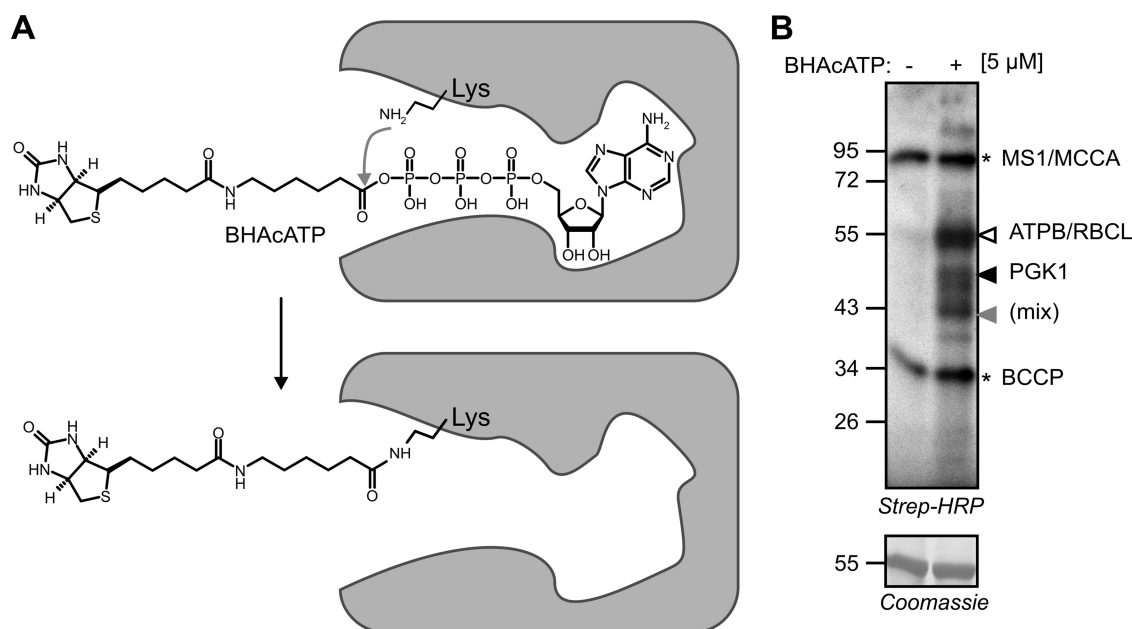
**Comparison with Protein Levels in the AtProteome Database**—For every labeled protein in supplemental Table S4, we retrieved the number of spectral counts from leaves from the AtProteome database (48). For ambiguous peptides, only the protein with the highest number of spectral counts was selected.

**In Silico Analysis of Labeling Sites**—To predict the labeled peptides for all the protein kinases in Arabidopsis, we retrieved protein sequences of all protein kinases from the Arabidopsis thaliana kinase database. Sequences for each subfamily in the AthKD database were retrieved and aligned using MultiAlign. The labeled peptide that was experimentally identified was highlighted in the alignment, the orthologous Lys residues were indicated, and Lys and Arg residues were highlighted in the alignment. Tryptic peptides containing the labeling site were selected from the alignments. The orthologous labeled Lys and the Arg/Lys-Pro sites were not considered as trypsin cleavage sites, and miscleavages were ignored. The selected peptides were ranked on sequence, and redundancy for each peptide sequence was counted (*n* proteins carrying the same labeled peptide). Ile/Leu residues were not discriminated in this analysis, as they have the same mass. Protein kinases with a unique labeled lysine (*n* = 1) were counted. The other protein kinases contain ambiguous peptides (*n* > 1). This analysis was done for peptides carrying Lys1, Lys2, Lys3, and Lys3.

## RESULTS

In this study, BHAcATP was used to label Arabidopsis leaf extracts. BHAcATP is composed of ATP, an acyl phosphate linker, and a biotin tag (Fig. 1A). Incubation of leaf proteomes with BHAcATP followed by the detection of protein blots with streptavidin-HRP revealed numerous biotinylated proteins when compared with the no-probe control (Fig. 1B). Signals at 55, 45, and 43 kDa were hallmark signals (Fig. 1B; white, black, and gray arrowheads, respectively) later identified as ATPB/RBCL, PGK1, and a mixture of ATP binding proteins, respectively (see below). The endogenously biotinylated proteins 3-methylcrotonyl-CoA carboxylase (MCCA) and biotin carboxyl carrier protein (BCCP) appeared as background signals at 80 kDa and 34 kDa, respectively, in both the labeled





**FIG. 1. Structure and mechanism of labeling with BHAcATP.** **A**, BHAcATP contains ATP, an acyl phosphate reactive group, and a biotin tag. When BHAcATP binds to the ATP binding pocket of a protein, the amino group of the nearby lysine reacts with the carbonyl carbon, which results in the covalent binding of the biotin tag to the protein while ATP is released. **B**, typical BHAcATP labeling profile of Arabidopsis leaf proteome. Arabidopsis leaf extracts were labeled with BHAcATP and the biotinylated proteins were detected on protein blots using streptavidin-HRP. Coomassie Brilliant Blue staining indicates equal loading. Asterisks indicate endogenously biotinylated proteins MCCA and BCCP. White, black, and gray arrowheads indicate bands containing ATPB+RBCL, PGK1, and a mix of ATP binding proteins, respectively. Abbreviations: MCCA, 3-methylcrotonyl-CoA carboxylase; BCCP, biotin carboxyl carrier protein; ATPB, chloroplastic ATPase; RBCL, ribulose-bisphosphate carboxylase; PGK1, phosphoglycerate kinase-1.

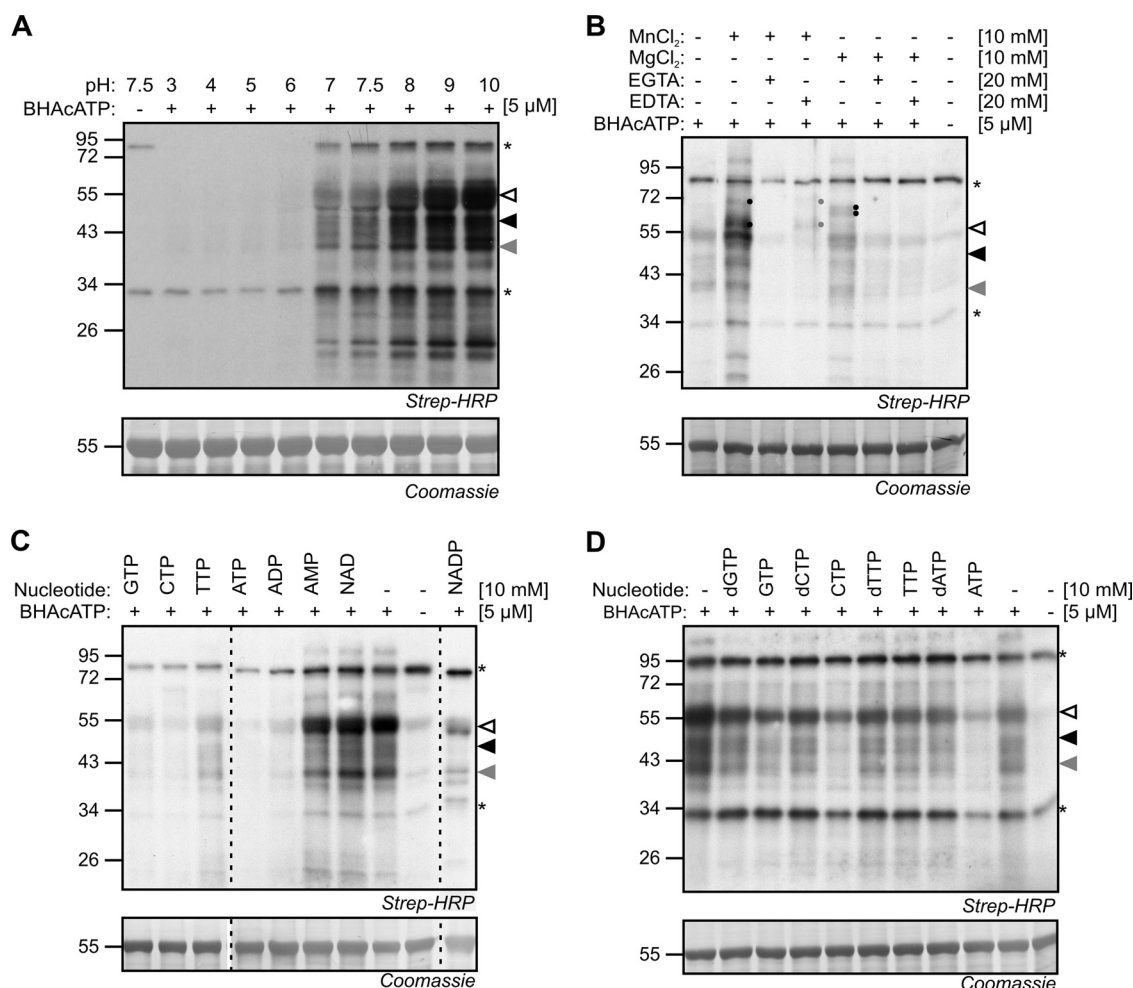
samples and the no-probe control. However, both these proteins are also ATP binding proteins, and the MCCA signal is mixed with methionine synthase, an ATP binding protein (see below).

**Labeling Is Dependent on pH and Divalent Metal Ions—** Different labeling conditions were tested in order to further characterize BHAcATP labeling. Labeling with BHAcATP depends on pH. Weak or no labeling occurs at acidic pH (pH 3–5), and strong labeling occurs at neutral to basic pH (pH 7–10; Fig. 2A). We chose pH 7.5 as the standard labeling condition to mimic the conditions of the plant cytoplasm. The addition of  $MgCl_2$  or  $MnCl_2$  greatly increased labeling, whereas adding the chelating agents EDTA and EGTA decreased labeling (Fig. 2B). Interestingly, additional 60- and 70-kDa signals appeared when  $MnCl_2$  was added, and two 65-kDa signals appeared when  $MgCl_2$  was added (Fig. 2B, solid circles). The  $MnCl_2$ -induced signals were suppressed more effectively by the addition of EGTA than of EDTA, whereas the  $MgCl_2$ -induced signals were suppressed equally by both EDTA and EGTA. These results are in agreement with the fact that EGTA has a higher affinity for  $Mn^{2+}$  than EDTA. The increased labeling upon the addition of divalent metal cations and the corresponding decrease upon treatment with chelating agents are consistent with the role of divalent ions in ATP binding.

**Labeling Is Suppressed with ATP-like Nucleotides—** We next tested whether labeling would be suppressed by nucle-

otides. Pre-incubation with ATP or ADP strongly suppressed BHAcATP labeling, whereas AMP did not suppress labeling (Fig. 2C), suggesting that phosphates at the  $\beta$  and  $\gamma$  positions are essential for the suppression of BHAcATP labeling. In addition to ATP, CTP and, to a lesser extent, GTP and TTP also suppress BHAcATP labeling (Fig. 2C). Notably, NADP, but not NAD, also suppresses BHAcATP labeling (Fig. 2C), indicating the importance of a phosphate moiety at the 2' position in the nucleotide to suppress labeling (supplemental Fig. S1). Interestingly, nucleotide triphosphates (NTPs) were more potent inhibitors than their deoxynucleotide counterparts (dNTPs; Fig. 2D), suggesting that the 2' hydroxyl group on the sugar moiety of the nucleotide also plays a role in determining the ability to compete for labeling (supplemental Fig. S1). In conclusion, these data indicate that labeling is specific because it can be competed with ATP analogues.

**BHAcATP Labels ATP Binding Proteins—** To identify the proteins labeled by BHAcATP, the labeled proteins were affinity purified using streptavidin beads, separated on protein gel, and stained (Fig. 3A). Bands were excised, treated with trypsin, and analyzed via LC-MS/MS. Peptides with high Mascot scores ( $>41$ ) and proteins with at least two unique peptides were retained. A total of 112 proteins were identified in the BHAcATP-labeled sample, but only 13 of these proteins were also identified in the no-probe control (supplemental Tables S1 and S2). The no-probe control contained proteins

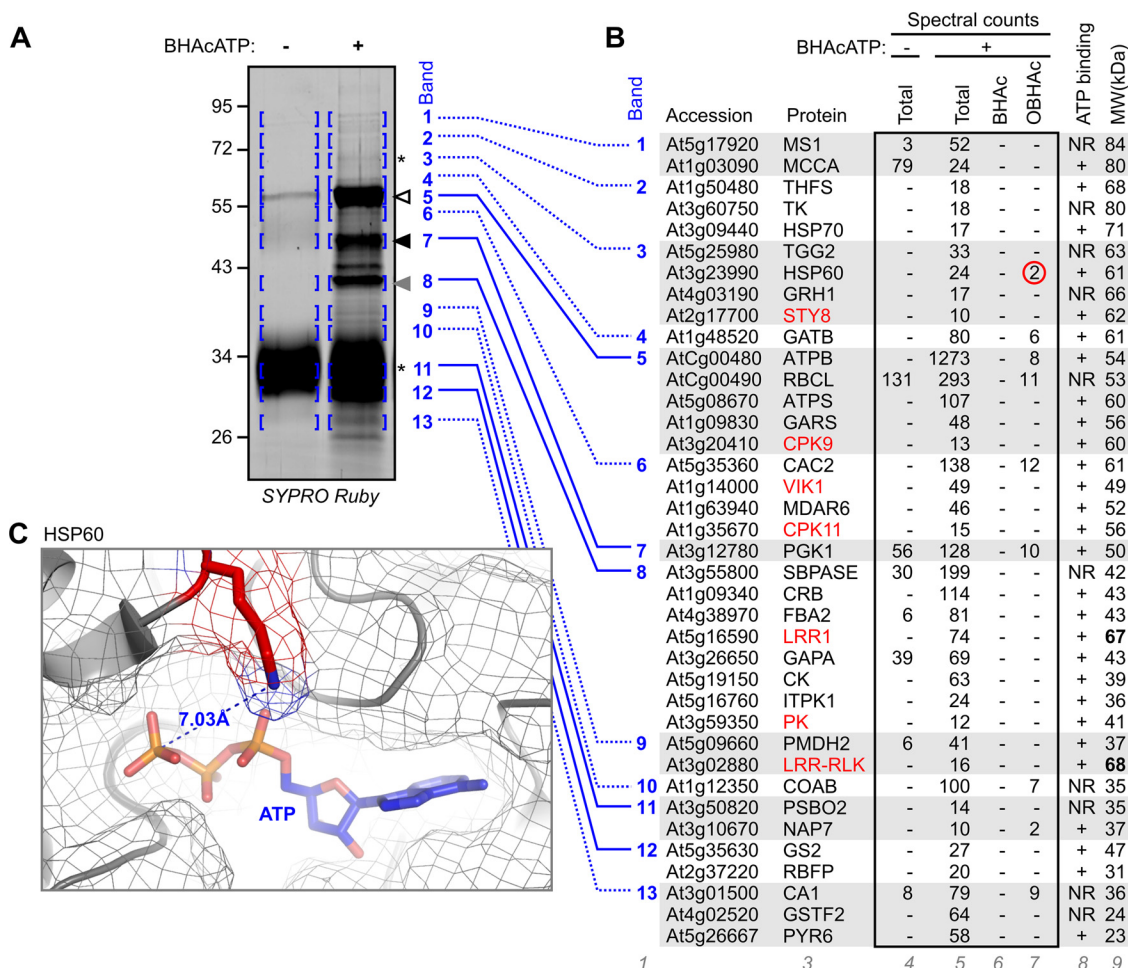


**FIG. 2. BHAcATP labeling depends on conditions.** A, BHAcATP labeling is dependent on pH. Arabidopsis leaf extracts were labeled with BHAcATP at various pH values, and biotinylated proteins were detected from protein blots using streptavidin-HRP. B, labeling is affected by MnCl<sub>2</sub> and MgCl<sub>2</sub>. Arabidopsis leaf extracts were labeled with 5 μM BHAcATP in the presence (+) or absence (-) of 10 mM MnCl<sub>2</sub> or MgCl<sub>2</sub> and/or 20 mM EDTA or EGTA. Biotinylated proteins were detected on protein blots using streptavidin-HRP. MnCl<sub>2</sub>- and MgCl<sub>2</sub>-induced signals (solid circles) are suppressed by EDTA and/or EGTA. C, BHAcATP labeling is outcompeted with a variety of nucleotide compounds such as GTP, CTP, TTP, ATP, ADP, and NADPH, but not with AMP and NAD. D, the nucleotide triphosphates (NTPs) GTP, CTP, TTP, and ATP were more efficient in competing labeling than their deoxynucleotide forms (dNTPs). For C and D, Arabidopsis leaf extracts were preincubated with nucleotide inhibitors for 30 min before labeling with 5 μM BHAcATP. Asterisks and arrows are as described in Fig. 1.

such as the endogenously biotinylated protein MCCA and abundant proteins such as RBCL.

We made a selection of the 112 proteins by selecting proteins with the highest spectral counts, supplemented with detected protein kinases and proteins for which a labeling site was identified (see below). The resulting selection consisted of 38 proteins that could be assigned to the 13 different bands (supplemental Table S2 and Fig. 3B). Some of these proteins also were assigned to the bands in Fig. 3A. These data indicate that the abundant 55-kDa signal (band 5) was caused by labeling of a subunit of the chloroplastic ATPase (ATPB) and the large subunit of ribulose-1,5-bisphosphate carboxylase oxygenase (RBCL). This signal also contained peptides from the beta subunit of the mitochondrial ATP synthase, glycylamide ribonucleotide synthetase, and CPK9. The 45-

kDa signal (band 7) was caused predominantly by phosphoglycerate kinase (PGK1), though this signal also contained spectra from other proteins. The 40-kDa signal (band 8) contained a mixture of chloroplastic sedoheptulose-1,7-bisphosphatase; chloroplast RNA binding protein; fructose-bisphosphate aldolase-2; subunit A of glyceraldehyde 3-phosphate dehydrogenase; and several other proteins, including two protein kinases (LRR1 and PK), inositol-1,3,4-trisphosphate-5,6-kinase, and an uncharacterized carbohydrate kinase. Weaker signals in the upper 70- to 90-kDa region of the gel contained methionine synthase (band 1), MCCA (band 1), formyltetrahydrofolate synthetase (band 2), transketolase (band 2), heat shock proteins 70 (HSP70, band 2) and 60 (HSP60, band 3), thioglucoside glucosylhydrolase 2 (band 3), F-box protein GRH1 (band 3), Ser/Thr/Tyr kinase STY8 (band

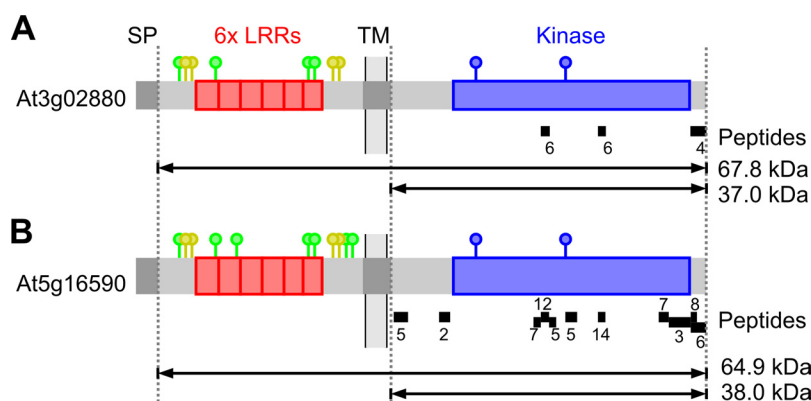


**FIG. 3. MS analysis of BHAcATP-labeled proteins.** A, Arabidopsis leaf extracts were incubated with and without 10  $\mu$ M BHAcATP and biotinylated proteins were purified, separated on protein gel, and detected using SYPRO Ruby staining. Thirteen bands were excised from each lane and treated with trypsin, and peptides were eluted and analyzed via LC-MS/MS. Identified proteins are indicated in B. See supplemental Tables S1 and S2 for more details. B, spectral counts of proteins in each of the 13 bands. Summarized are band numbers (column 1), accession codes (column 2), protein names (column 3), spectral counts in no-probe control (column 4), and the AcATP-labeled sample (column 5). The spectral counts also include BHAc- and OBHAc-modified peptides, indicated separately in columns 6 and 7. Expected ATP binding based on literature data is indicated in column 8 (ATP binding activity reported (+) or not reported (NR)). The theoretical MW for each protein is given in column 9. Deviations from the actual MW are printed in red. This table is a selection of the most abundant spectral counts per band, supplemented with detected protein kinases and proteins for which labeled peptides were detected. Proteins that were detected in more than one band are shown only in the band with the highest spectral counts. For a complete list of identified proteins, see supplemental Tables S1 and S2. C, AcATP labeling site in HSP60. The orthologous labeling site of AcATP in HSP60 of Arabidopsis is indicated in the crystal structure of the highly homologous GroEL of *E. coli* (1sx3, 16) containing ATP. The labeled lysine was located from the aligned protein sequences and highlighted in the structure using PyMol. The lysine is 7.03 Å from the gamma phosphate of ATP.

3), and subunit B of glutamyl-tRNA aminotransferase (band 4). Weak signals at 50 kDa (band 6) contained a subunit of a biotin carboxylase (CAC2), a MAP3K protein kinase (VIK1), a monodehydroascorbate reductase (MDAR6), and CPK11. Weak signals in the 35- to 40-kDa region contained peroxisomal NAD-malate dehydrogenase-2 (band 9), an LRR-RLK (band 9), and phosphopantothenate-cysteine ligase (COAB, band 10). The signal at 33 kDa (band 11) contained predominantly avidin, but also present were subunit O<sub>2</sub> of photosystem II and non-intrinsic ABC protein-7. Finally, the weak signals at the bottom of the gel (27–30 kDa) contained chloroplastic glutamine synthetase (GS2, band 11), chloro-

plast RNA-binding protein (band 12), carbonic anhydrase (band 12), glutathione transferase (band 12), and a UMP/CMP kinase (PYR6, band 12).

In conclusion, the vast majority of the identified proteins in the AcATP-labeled samples are ATP binding proteins (Fig. 3B). These ATP binding activities include a variety of protein kinases (STY8, CPK9, VIK1, CPK11, RLK-LRR, and LRR1), ATP-based transporters (ATPB, ATP synthase, and non-intrinsic ABC protein-7), metabolite kinases (carbohydrate kinase, PGK1, inositol-1,3,4-trisphosphate-5,6-kinase, and PYR6), and metabolic enzymes and chaperones using ATP (glycinamide ribonucleotide synthetase, MCCA, formyltetra-



**FIG. 4. Putative receptor shedding of two receptor-like kinases.** RLKs At3g02880 and At5g16590 both carry a signal peptide (SP), an N-terminal LRR-flanking domain with a conserved Cys pair (yellow), six leucine-rich repeats (LRRs; red), an extracellular juxtamembrane domain with a conserved Cys pair (yellow), a transmembrane domain (TM), an intracellular juxtamembrane domain, a protein kinase domain (blue), and a C-terminal tail. The extracellular domain carries several putative N-glycosylation sites (green). AcATP labeling sites detected via gel-free analysis (Fig. 7) are indicated in blue. Identified peptides and spectral counts are indicated with black bars below the proteins. The theoretical molecular weight was calculated from the plain amino acid sequence for the cytoplasmic domain and the full mature protein without signal peptide.

hydrofolate synthetase, HSP70, HSP60, CAC2, MDAR6, peroxisomal NAD-malate dehydrogenase-2, and GS2).

**Putative Receptor Shedding of RLKs**—We next examined whether the experimental molecular weight (MW) of the proteins identified from the protein gel (Fig. 3B) corresponded to the theoretical MW calculated from the protein sequence. We noticed a good correlation for nearly all the detected proteins, with the striking exception of two LRR-RLKs that had a calculated MW of 68 kDa but migrated in the region of 40 kDa (Fig. 3B). These were two highly homologous LRR-RLKs carrying six extracellular LRRs, a transmembrane domain, and a cytoplasmic kinase domain (Fig. 4). The extracellular domain also contains multiple putative N-glycosylation sites, which causes LRR-RLK proteins to migrate with an apparent MW that is typically 20 to 30 kDa larger than their calculated MW. Peptide coverage (Fig. 4 and supplemental Table S1) suggests that these labeled proteins consist of the full protein kinase domain but lack the extracellular domain. The detection of half-tryptic peptides of the C-termini of both these LRR-RLK proteins indicates that the C-terminus is intact and that a significant truncation must have occurred from the N-terminus. Interestingly, the calculated MW of the cytoplasmic domain is 37 and 38 kDa for At3g02880 and At5g16590, respectively, which coincides with the observed MW of these proteins. Taken together, these data indicate that these LRR-RLKs exist in leaves as kinase domains lacking the extracellular domain.

**Detection of an Oxidized Biotin Moiety**—The searches mentioned above included a variable modification of lysine residues with the probe and allowed two trypsin miscleavages. Searches with the theoretical BHAc modification on lysine (339.16 Da) did not lead to hits with Mascot scores above the threshold (Fig. 3B). However, during more liberal searches we noticed that BHAc modifications appeared, but only in pep-

tides that included an oxidation associated with the presence of a methionine in the peptide sequence (data not shown). This observation sparked the idea that the modification of the probe is associated with an additional oxygen. Indeed, when we searched the data with an oxygen added to the BHAc modification (OBHAc, 355.16 Da), we were able to identify modified peptides for nine different proteins with relatively high Mascot scores (41–71) (Fig. 3B).

One of those labeled peptides is from HSP60 (At3g23990), a well-described ATP binding protein. To determine the location of the labeling site in the protein relative to ATP, we searched the protein database for proteins homologous to HSP60 and identified GroEL, an HSP60 protein from *Escherichia coli*, for which an ATP-bound crystal structure is available (1sx3 (16)). HSP60 and GroEL proteins share 56% amino acid identity, including the region where the labeled lysine is located. The amino group of the labeled lysine is located 7.0 Å from the gamma phosphate of the bound ATP in GroEL (Fig. 3C). Thus, labeling in HSP60 occurred at the expected labeling site.

We next studied the fragmentation spectrum of the labeled peptide of HSP60 further to identify the location of the oxygen within the labeled peptide. The labeled peptide in HSP60 has the sequence VTKDGVTVAK and was detected with a Mascot score of 41.4 with a parental ion of 686.93  $m/z$  ( $z = 2$ ) (Fig. 5). Nearly all b- and y-ions were detected in the fragmentation spectrum, locating the labeled lysine including the oxygen at the third position in the peptide: VTK<sup>+</sup>DGVTVAK. The spectrum also contained three unexplained peaks in the low-mass range: 243.15, 356.31, and 439.35 Da. Importantly, the 243.15 Da ion corresponds to an oxidized biotin (expected mass: 243.3). An additional 113 Da for the aminohexanoic acid linker predicts an ion of 356.5 Da, which corresponds to the second signal at 356.31 Da. The third ion, observed 83 Da further at



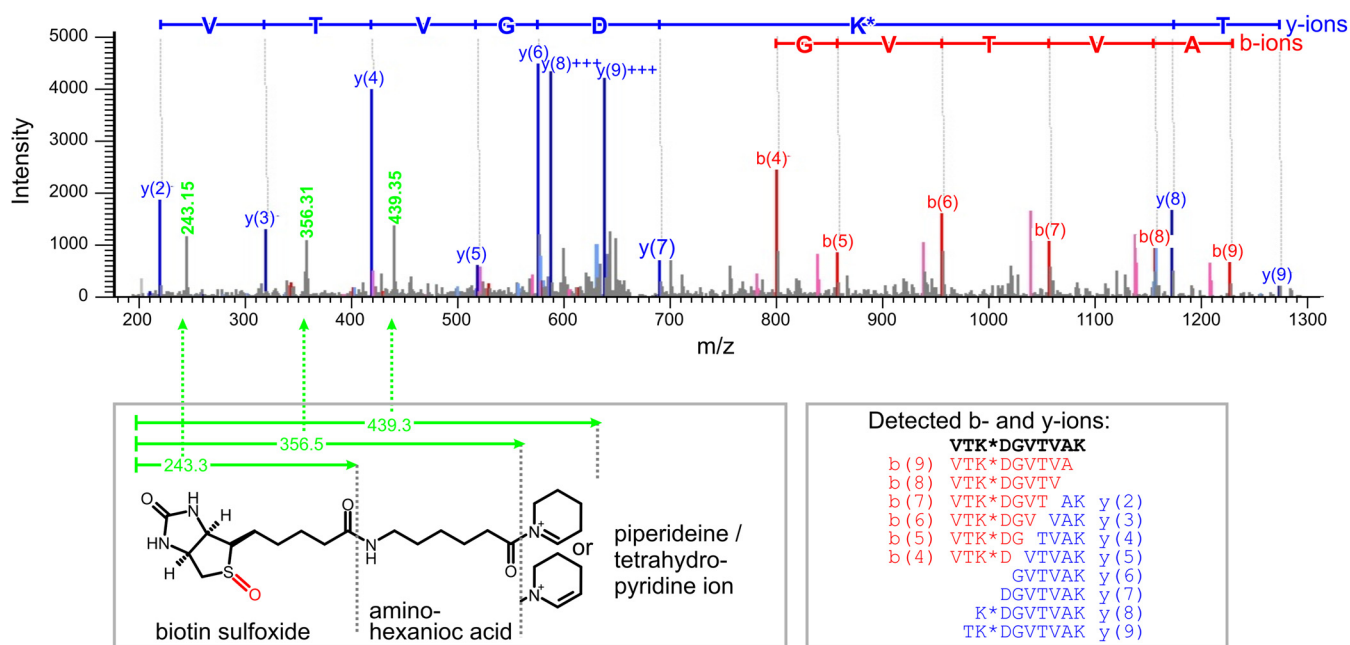


FIG. 5. **Fragmentation spectrum of HSP60 peptide labeled with oxidized biotin.** The b- and y-ion series is annotated as explained in the bottom right. The putative structure of the modified lysine is shown on the bottom left, explaining the three observed ions. The oxide (red) is presumably on the sulfur of the biotin, resulting in biotin sulfoxide.

439.35 Da, corresponds to the labeled version of the commonly observed immonium ion variant for lysine. High-energy collision-induced dissociation fragmentation of lysine residues typically results in an unsaturated cyclic 84 Da fragment, known as a piperidine or tetrahydropyridine ion (17). As the lysine has lost a proton during the labeling reaction, the observed 83 Da distance is explained, indicating that such fragmentation and subsequent cyclization of lysine do not cause a loss of the biotin. Taken together, these data demonstrate that the additional oxygen is located at the biotin, presumably on the sulfur. The 243.2, 356.5, and 439.3 Da ions were also found in spectra of the other labeled peptides, indicating that oxygen is consistently present on the biotin moiety.

**Xsite Identifies 242 Labeling Sites**—The observation that most of the 112 BHAcATP-labeled proteins are ATP binding proteins in a crude proteome. However, only a few of these 112 proteins are protein kinases. Furthermore, some of the BHAcATP-labeled proteins are not annotated as ATP binding proteins, and therefore the labeling sites on these proteins are uncertain. To address both issues, we analyzed the labeled proteome more deeply by analyzing the labeled peptides. For this we used a variant of the AcATP probe, DBAcATP (Fig. 6A) (7). Desthiobiotin cannot be oxidized and has a lower affinity for streptavidin, resulting in a more efficient elution of peptides for identification. After labeling with DBAcATP, the proteomes were digested with trypsin, and the biotinylated peptides were purified using streptavidin beads and analyzed via LC-MS/MS (Fig. 6B). We recorded 10,465 peptide spectra, corresponding to 567 different peptides carrying a modified

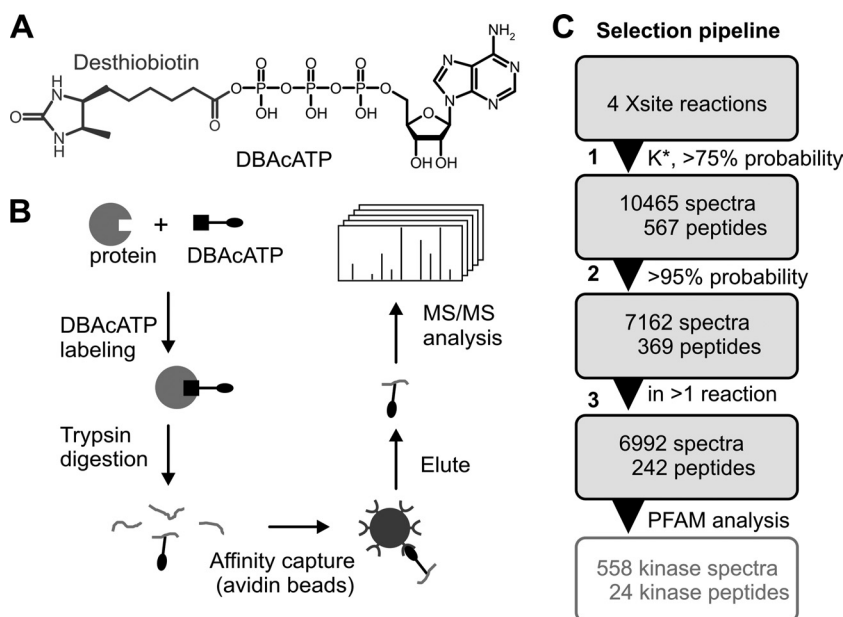
lysine residue (Fig. 6C). Peptides with probability scores greater than 95% (15) and those that were identified in two or more reactions were selected (Fig. 6C), resulting in 6992 spectra corresponding to 242 peptides (supplemental Table S4).

**DBAcATP Labels a Diversity of Protein Kinases**—We first analyzed the data for peptides derived from protein kinases. Annotation through the PFAM database revealed that 24 of the 242 labeled peptides originated from at least 21 different protein kinases (Fig. 7A; supplemental Table S4). This list contains seven RLKs, of which six carry extracellular LRRs. The detection of RLKs is remarkable, because we did not enrich for membranes. In addition, we identified protein kinase peptides from several MPKs, CPKs, PTI-like kinases, and other Ser/Thr protein kinases. We also detected AvrPphB susceptible 1 (PBS1) (18), VH1-interacting kinase (VIK) (19), AT6 protein kinase (20), and proline-rich extension-like receptor kinase-1 (PERK1) (21). Of the 24 protein kinase peptides, 19 were unambiguous and were derived from only one protein kinase (black accession numbers in Fig. 7A). Four protein-kinase-derived peptides were ambiguous, as the peptide sequence was identical in two or more protein kinases in the TAIR10 protein database (gray accession numbers in Fig. 7A). Two protein-kinase-derived peptides were identified for two protein kinases (top in Fig. 7A), indicating that DBAcATP labels these protein kinases at two positions. In the case of PERK1, two overlapping peptides containing the same labeled lysine were detected.

We marked the identified protein kinases in the *Arabidopsis thaliana* kinase database, which classifies all the 1099 protein



**FIG. 6. Identification of labeled peptides by Xsite.** A, structure of desthiobiotin-acyl-ATP (DBAcATP). The desthiobiotin affinity tag is used because it cannot be oxidized and has a lower affinity to avidin, which facilitates efficient peptide elution. B, leaf extracts were labeled with DBAcATP and trypsin-digested, and the labeled peptides were purified and analyzed via LC-MS/MS. C, pipeline for analysis of Xsite reactions. A total of four Xsite reactions were analyzed. Peptides were selected for (i) having labeled lysines with probability scores greater than 75%, (ii) having >95% probability, and (iii) being found in at least two of the four Xsite reactions. To select peptides from PKs, PFAM analysis was performed on the list of labeled peptides, resulting in a list of 558 spectra corresponding to 24 different peptides from PKs.



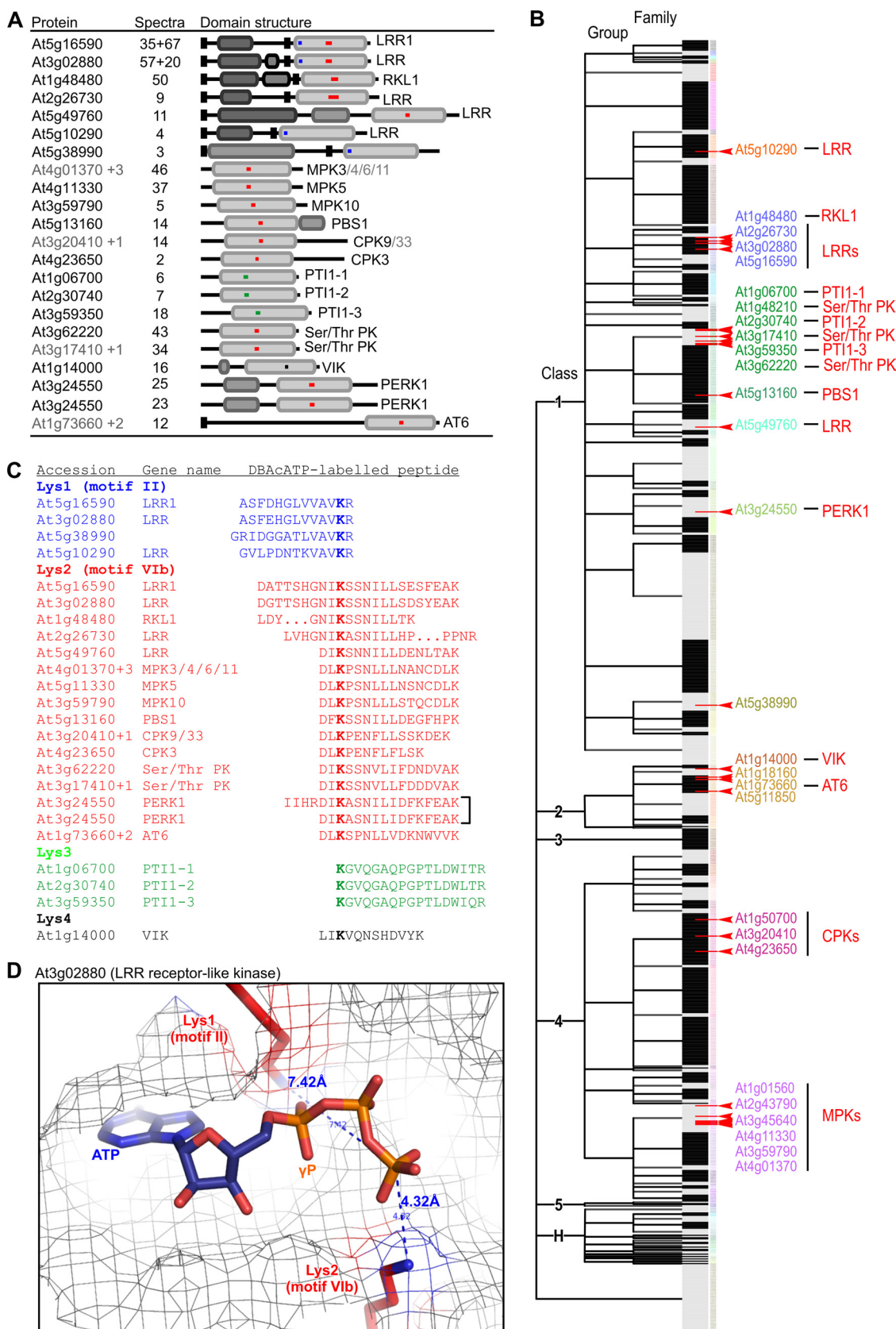
kinases of Arabidopsis into classes, groups, and families based on sequence similarities. We identified representatives of 11 different protein kinase families belonging to 10 different groups and the three major protein kinase classes (Fig. 7B). This indicates that AcATP probes are not family specific and can be used to profile ATP binding activities of the majority of Arabidopsis protein kinases.

Alignment of the identified protein-kinase-derived peptides showed that these peptides fall into four groups containing four different labeled lysine residues, Lys1–4 (Fig. 7C). When mapped on the protein kinase domains, Lys1 is positioned in the beginning of the protein kinase domain, Lys2 in the middle, Lys3 in the second quarter, and Lys4 in the third quarter of the protein kinase domain (Fig. 7A). Lys1 and Lys2 are also part of the conserved protein kinase motifs II and VIb, respectively (22). These two lysines are also present in the PTO kinase, a tomato protein kinase involved in immunity, for which a structure is available (2qkw) (23). Both Lys1 and Lys2 reside in the ATP binding pocket of the PTO kinase, at distances of 7.43 and 4.32 Å from the gamma phosphate of the bound ATP (Fig. 7D and supplemental Fig. S2A). The closer proximity of Lys2 seems reflected in the higher spectral count of Lys2-containing peptides relative to Lys1-containing peptides. Modeling of PTI1–1 indicates that Lys3 locates in an extended loop that might fold back on the ATP binding pocket (supplemental Fig. S2B), whereas modeling of VIK1 indicates that Lys4 is located in close proximity to the gamma phosphate of bound ATP (supplemental Fig. S2C).

To determine how common these labeled lysine residues are within the protein kinase family of Arabidopsis, we counted these lysines in the alignments of each protein kinase family. We found that Lys1 and Lys2 were conserved in 70% and 47% of the 1099 protein kinases, respectively. Lys3 in PTI

kinases and Lys4 in VIK1 were less conserved among protein kinases (3.5% and 5.0%, respectively). Thus, the frequencies of the theoretically labeled lysine in protein kinases correspond to the frequencies of the different labeling sites detected in protein kinases (supplemental Fig. S6).

**DBAcATP Also Labels Other ATP Binding Proteins in Their ATP Binding Pockets**—When ranked according to peptide frequency, peptides from protein kinases are not among the top 20 most frequent peptides (Fig. 8A). The most frequent peptide from a protein kinase is that from LRR1 (At5g16590) on position 22 with 67 spectral counts (Fig. 8A and supplemental Table S4). The most frequently detected labeled peptide is from PGK1, with 968 spectral counts. PGK1 is also labeled at two additional lysine residues with 196 and 97 spectral counts, respectively. The second most frequent labeled peptide is from chloroplast ATP synthase subunit beta (ATPB), with 256 spectral counts. Other proteins in the top 20 are acetyl coenzyme A (CAC2), subunit B of Glu-tRNA aminotransferase, an ATP synthase, ribose-phosphate pyrophosphokinase 4, the small subunit of ribulose biphosphate carboxylase, cell division cycle 48 (CDC48), a GTP binding protein (At1g30580), and UMP/CMP kinase (PYR6) (Fig. 8A). Most, but not all, of these proteins were also detected abundantly using the in-gel approach (Fig. 3). Although these proteins are unrelated and catalyze very different reactions, ATP is a common substrate, consistent with the reactivity of the probe (supplemental Table S5). Protein structures of homologs of PGK1, ATPB, CAC2, and ribose-phosphate pyrophosphokinase 4 contain the labeled Lys in close proximity to the terminal phosphate of the bound nucleotide (supplemental Fig. S3). Note that CAC2 is a biotin binding protein, but the labeled Lys is close to the bound ADP (supplemental Fig. S3C). This demonstrates that the labeling of these proteins is



in accordance with the labeling mechanism of AcATP probes. However, not all proteins in this list are known to bind nucleotides. Most evident is the large subunit of Rubisco (RBCL), which is labeled at 13 different lysine residues that are scattered on the surface of the RBCL protein (supplemental Fig. S4).

**Potential of DBAcATP Labeling in Arabidopsis Proteome**—We extended the analysis of labeling sites in homologous protein structures and found that 63 of the 242 labeled peptides resulted from labeling of a lysine residue in a known nucleotide binding pocket at  $\leq 10$  Å from the phosphate of the bound nucleotide (supplemental Fig. S5; blue in Fig. 8A). To estimate the potential of DBAcATP labeling in the Arabidopsis proteome, we retrieved for each of the 63 labeled peptides the corresponding PFAM domain. A total of 26 different PFAM domains corresponding to 13 different protein clans plus six uncategorized PFAM families were identified (supplemental Table S6). The Arabidopsis genome encodes for a total of 1683 proteins carrying these 26 different PFAM domains. The largest PFAM domains are those of protein kinases (PF00069 = 813 members and PF07714 = 388 members), followed by AAA ATPases (147 members). Because all 1683 of these proteins are probably binding ATP and representatives were found to be labeled by DBAcATP, we predict that the Arabidopsis proteome contains at least 1683 putative targets that could be labeled in the ATP binding pocket by DBAcATP.

**Preferential Labeling of ATP Binding Proteins**—To investigate whether protein kinases and other ATP binding proteins are preferentially labeled relative to highly abundant proteins, we plotted the spectral counts of the labeled peptides against the spectral counts detected for these proteins in leaf proteomes of the AtProteome database (Fig. 8B). We chose this approach to ensure the inclusion of low-abundance protein kinases, which were detected by Baerenfaller *et al.* (48) with over 1300 MS runs. The leaf proteomes that they used are virtually the same as the ones we used for our studies. Interestingly, this graph splits the labeled peptides into three groups. Peptides from protein kinases cluster in group 1 at low protein levels and medium labeling. Peptides from other ATP binding proteins are in group 2 and show medium protein levels and medium to high labeling. Finally, the third major group contains proteins with high protein levels and low to

medium labeling. Importantly, the vast majority of these proteins were not labeled in ATP binding pockets. This illustrates that DBAcATP shows a high selectivity for labeling ATP binding proteins, as only highly abundant proteins are labeled outside known ATP binding pockets.

Interestingly, protein kinases were detected at a moderate labeling/protein ratio relative to labeling events in ATP binding pockets of non-protein kinases ( $0.5 \pm 0.57$  ( $n = 15$ ) versus  $2.86 \pm 0.68$  ( $n = 34$ )) (Fig. 8C). These data suggest that the protein kinases have a lower affinity for ATP or that that part of the protein kinase pool cannot be labeled. In order to investigate this in more detail, we extracted spectral counts for 494 protein kinases from leaf proteomes of the AtProteome database (48). Interestingly, of the top 20 protein kinases that are most frequently detected in leaf proteomes, only 8 were detected upon labeling (supplemental Fig. S7), and these data show no correlation between spectral counts in the AtProteome database and spectral count frequency upon AcATP labeling. Furthermore, inspection of the protein kinase sequences revealed that most of the undetected labeled peptides fell in a MW range that should have been detected in our assay. These data reinforce the idea that not all kinases can be labeled in a given extract, possibly because they do not bind ATP. However, such data should be interpreted with extreme caution, because different peptides have different ionization potentials, and the absence of a labeled peptide does not mean that the protein is not labeled.

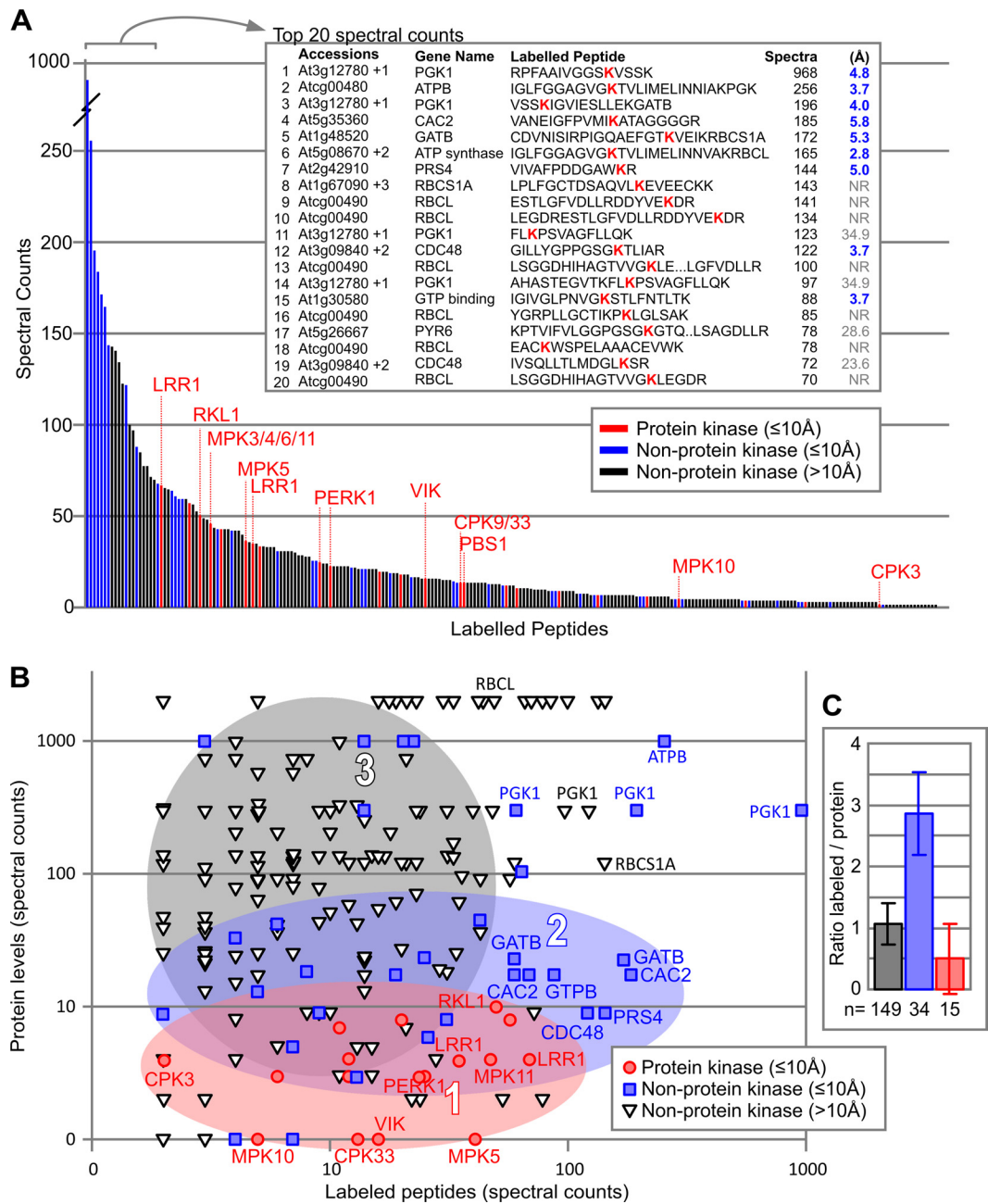
## DISCUSSION

We have reported the first in-depth analysis of targets of an AcATP probe in a plant proteome. By comparing a gel-based identification platform for labeled proteins with a gel-free platform for labeled peptides, we have demonstrated the advantages and limitations of these complementary approaches. The analysis of labeling sites using the protein database showed that AcATP probes target predominantly, but not exclusively, the ATP binding pockets of a broad range of protein kinases and other unrelated ATP binding proteins. The labeled proteins include a few well-characterized ATP binding proteins, as well as many ATP binding proteins that have not been studied so far.

**A Comparison of Two Analysis Platforms**—The data generated by the gel-based and gel-free platforms are complemen-

**FIG. 7. Protein kinases labeled with DBAcATP.** A, protein domains are indicated as boxes, with protein kinase domains containing bars representing identified labeled peptides containing Lys1 (blue), Lys2 (red), Lys3 (green), and Lys4 (black). Protein kinases that match the same peptides are indicated with gray accession numbers and summarized as At4g01370 + 3 = At4g01370, At3g45640, At2g43790, and/or At1g01560; At3g20410 + 1 = At3g20410 and/or At1g50700; At3g17410 + 1 = At3g17410 and/or At1g48210; and At1g73660 + 2 = At1g73660, At5g11850, and/or At1g18160. Common names of protein kinases are indicated, as well as corresponding spectral counts. B, classification of the Arabidopsis protein kinases according to the *Arabidopsis thaliana* Kinase Database (AthKD). The identified labeled protein kinases are indicated with red arrowheads. Accession numbers and common names are shown. Detected protein kinases represent 11 families that belong to 10 groups and 3 classes. C, alignment of identified labeled peptides containing four different labeled lysines. Lys1 and Lys2 belong to protein kinase motifs II and VIb, respectively. Color coding is the same as in A. Labeled lysine residues are printed in bold. D, position of the two conserved labeled lysine residues in the ATP binding pocket of LRR-RLK protein At3g02880, based on the crystal structure of the highly homologous protein kinase PTO (23). The distances from the labeled Lys to the gamma phosphate of the bound ATP molecule are shown.





**FIG. 8. Spectral count analysis of labeled peptides.** A, labeled peptides ranked on spectral count frequencies. Labeling sites in the ATP binding pocket of protein kinases and other nucleotide binding proteins are indicated in red and blue, respectively. The top 20 peptides with the highest spectral counts are shown in the box, along with their accession numbers, their common names, the peptide sequences (labeled Lys, bold), and their spectral counts. Distances from the nucleotides to the labeled Lys are indicated in Å. NR, not reported nucleotide binding site. Ambiguous peptides are summarized as follows: At3g12780 + 1 = At3g12780 and/or At1g56190; At5g08670 + 2 = At5g08670, At5g08680, and/or At5g08690; At1g67090 + 3 = At1g67090, At5g38430, At5g38420, and/or At5g38410; At3g09840 + 2 = At3g09840, At3g53230, and/or At5g03340. B, correlations between ATB binding activity and protein abundance splits the labeled peptides into three groups. The frequency of spectral counts of the labeled peptides is plotted against the spectral counts of the corresponding proteins detected in leaf proteomes (extracted from the AtProteome database). For ambiguous labeled peptides, the protein with the highest spectral count was selected from the AtProteome database. Note that for some proteins (e.g. RBCL) multiple labeled peptides were detected. Group numbers 1–3 are explained in the main text. C, average spectral count ratio between labeling (by AcATP) and abundance (extracted from AtProteome database) for the three groups of labeled proteins. Error bars represent S.E. for *n* proteins (indicated below the bars).

tary, and each platform has its advantages and disadvantages. Although many proteins were detected using both approaches, several were detected by only one approach. HSP60 and HSP70, for example, were detected only using the in-gel approach, whereas CDC48 and MPKs were detected only via the gel-free approach. Gel-based identification resulted in the detection of 112 proteins, of which only 9 were protein kinases. Although the vast majority of the detected proteins were absent in the no-probe control, the labeling site could be determined for only nine proteins, based on the detection of labeled peptides. This number would have been even lower if we had not realized that biotin is somehow oxidized during this procedure. By contrast, sequencing the labeled peptides using Xsite/KiNativ via the gel-free approach identified more protein kinases and also determined labeling sites in each of the labeled proteins. Although this gel-free approach seems more powerful, it cannot discriminate between proteins if the labeled peptide is identical. For example, the labeled peptides do not discriminate between CPK9 and CPK11 because the labeled peptide is identical, whereas both proteins were distinguished via the gel-based approach. The gel-free approach also does not provide the MW of the labeled protein, and this can be very important information. For example, we noticed that two labeled LRR-RLKs migrated not at the expected 67–68 kDa, but at 40 kDa, indicating that they had lost the extracellular LRR domain. At this stage we cannot exclude the possibility that this processing occurred upon protein extraction. It is interesting to point out that it was recently discovered that Xa21, an LRR-RLK from rice, is proteolytically cleaved during signaling to release a cytosolic protein kinase domain that migrates to the nucleus to phosphorylate transcription factors (24). A similar receptor-shedding mechanism might be at work for the two LRR-RLKs detected on our study.

**Detection of an Oxidized Biotin Moiety**—Our analysis of biotinylated proteins via in-gel digest and LC-MS/MS revealed no peptides that were labeled with BHAc but several peptides that were labeled with OBHAc, the oxidized version of BHAc. To our knowledge, this type of modified biotin has not been reported before during proteomics, and this discovery might have important implications for the detection of other modifications containing biotin. We speculate that the oxygen is located on the sulfur of biotin, resulting in a biotin sulfoxide (25). This modification was also observed, for example, during the synthesis of oligonucleotides (26), and would not occur on desthiobiotin. At this stage we do not know when during the in-gel procedure the oxidation occurs. Searches on the occurrence of oxidized biotin on proteins labeled with other probes will tell how common this modification is.

**How to Detect More Protein Kinases?**—We detected 24 labeled peptides from protein kinases, but there are 1099 protein kinases encoded by the Arabidopsis genome (1). There are several reasons why no more protein kinases were

detected. First, most of the protein kinases are not expressed in tissues that we used for the labeling experiments. Second, peptides from some protein kinases are too small or too large to be detected with the chosen MS settings. However, *in silico* analysis indicates that the majority (95%) of the theoretically labeled peptides of Arabidopsis protein kinases are in the range of 5 to 30 amino acids. Third, some protein kinases cannot be discriminated based on the sequence of the labeled peptides. Nonetheless, *in silico* analysis showed that 64% of the Arabidopsis protein kinases contain a putative labeled lysine in a peptide sequence that is unique in the Arabidopsis proteome (supplemental Fig. S6). Fourth, MS is not sensitive enough to detect more labeled peptides from protein kinases in an unbiased mode. Indeed, inclusion lists coupled to retention times and searches for marker ions in the MS2 mode can tremendously increase the number of identified protein kinases. The added search modes have increased the number of robustly detected protein kinases to 160 protein kinases per cell line (7). Thus, we expect that we will be able to detect labeling of over 100 different protein kinases from a single proteome with the implementation of empirical searches. The wider range of detection will increase the power of protein kinase profiling in plants tremendously.

**Profiling Arabidopsis Protein Kinases**—We have detected 24 labeled peptides that originate from protein kinases. Analysis of the labeling sites in these protein kinases showed that AcATP probes react with lysines that reside at four distinct positions. Lys1 and Lys2 reside in the ATP binding pocket in motifs II and VIb, respectively, and were previously described to be labeled by AcATP probes (6). Lys3 and Lys4 were detected in PTI1-like kinases and MAP3K VIK, respectively, and had not been characterized before. Although crystal structures are not available for representatives of these protein kinase subfamilies, modeling indicates that regions carrying Lys3 and Lys4 might indeed be able to fold back on the ATP binding pocket. The reactivity of Lys3 and Lys4 indicates that even though these lysines are poorly conserved, they might be in close proximity to ATP and might play important roles in PTI1- and VIK-like protein kinases. The absence of any other labeled peptide from the detected protein kinases illustrates the remarkable selectivity of the AcATP probe to target the ATP binding pocket of these proteins.

The labeled protein kinases are remarkably diverse and contain representatives of all major classes of Arabidopsis protein kinases. The list includes RLKs, Pelle-like kinases, MAPKs, MAP3Ks, CPKs, PTI-like kinases, and PERK1. The diversity of the protein kinases illustrates the wide range of protein kinases that can be labeled with AcATP and is consistent with the fact that all protein kinases (by definition) bind ATP and most of them carry lysine residues to stabilize the phosphate. The data are also consistent with studies of AcATP probes on animal proteomes, where >75% of the human protein kinases have been detected (6). Thus, AcATP probes have the remarkable potential to label nearly all protein

kinases not only of mammals, but also of other organisms. The detection of several RLKs with relatively high spectral counts is noteworthy because these proteins are thought to be low in abundance and we did not enrich for membrane proteins.

Functions have been described for several of the detected protein kinases. MPK3, -4, -6, and -11 are involved in immune signaling and control gene expression through WRKY transcription factors (4, 27, 28). PBS1 is involved in the perception of bacterial pathogens secreting the AvrPphB effector (29, 30). AvrPphB cleaves PBS1, and this activates resistance protein RPS5. Calcium-dependent kinase CPK3 is involved in stomatal closure, herbivore defense, and salt stress acclimation and is a positive regulator of sphingolipid-induced cell death (31–33). PTI1–2 interacts with oxidative stress-response protein kinase OXI1 and integrates phosphatidic acid signaling with reactive oxygen signaling (34). VIK is a MAP3K involved in the uptake of glucose into the vacuole (19), and AT6 is a MAP3K that acts as a negative regulator of salt tolerance (20). PERK1 is thought to be involved in sensing cell wall stress during wounding and pathogen infection (21, 35). The detection of these characterized protein kinases via AcATP labeling is consistent with their presence and activities in the Arabidopsis rosette used for our studies. The other two-thirds of the detected protein kinases have been described only by transcript levels and phylogeny. Our data demonstrate that these uncharacterized protein kinases are present and able to bind ATP.

**Labeling of Other ATP Binding Proteins**—Previous studies with AcATP probes have been mostly focused on the labeling of protein kinases. In this study we also examined the labeling of proteins other than protein kinases. An in-depth analysis of the labeled peptides revealed that AcATP probes label an astonishing diversity of protein families. Most of the labeled proteins are able to bind nucleotides, and analysis of structural data indicates that AcATP probes label these nucleotide binding proteins predominantly inside ATP binding pockets. The structural diversity of the nucleotide binding proteins is remarkable, as illustrated by the fact that they belong to 13 different protein clans in the PFAM database (36).

Several different ATP binding proteins have been identified with high spectral count frequencies. PGK1 is the top hit among the labeled proteins. PGK degrades ATP during glycolysis (37). Subunit B of glutamyl-tRNA aminotransferase is an Asn/Gln-tRNA amidotransferase involved in the transamination of misacylated Asp-tRNA or Glu-tRNA forming a correctly charged Asn-tRNA or Gln-tRNA, a reaction that requires ATP (38, 39). CDC48 encodes for cell division cycle protein 48 and is important in cell division, expansion, and differentiation. Arabidopsis plants carrying mutations in CDC48-encoding genes show impaired seedling development and defective pollen and embryo development (40). CDC48 is an AAA-type ATPase and belongs to a large superfamily containing a highly conserved triple-A domain that binds and hydrolyzes ATP

(41). CAC2 is the biotin carboxylase subunit of the heterodimeric, biotin-containing enzyme chloroplastic acetyl-coenzyme A carboxylase, which is involved in the synthesis of fatty acids in Arabidopsis (42) and uses ATP for the conjugation of a carboxylate-containing molecule to an amino- or thiol-group-containing molecule (43). CAC2 contains a phosphate-binding loop and a  $Mg^{2+}$ -binding site and has two alpha-beta subdomains that grasp the ATP molecule. This “ATP-grasp” domain is common to a superfamily that also contains D-alanine:D-alanine ligase, glutathione synthetase, and carbamoyl phosphate synthase. PYR6 is a uridine 5'-monophosphate (UMP)/cytidine 5'-monophosphate (CMP) kinase that converts UMP/CMPs to uridine and cytidine diphosphates, respectively. PYR6 orthologs in bacteria and yeast are required for cellular proliferation and RNA and protein synthesis (44, 45). PYR6 has also been detected in the mature pollen of Arabidopsis, consistent with its role in cell proliferation and division (46). Taken together, these five structurally and functionally diverse examples illustrate the broad range and relevance of ATP binding enzymes that we can monitor using AcATP probes.

**Labeling Outside Known ATP Binding Pockets**—Our analysis of labeling sites also showed that a large number of labeling sites are not located in known nucleotide binding pockets. It is possible that several of these labeling sites are nucleotide binding sites that have not yet been described. However, the correlation between spectral counts and transcript levels indicates that labeling outside known nucleotide binding pockets is typically detected for highly abundant proteins, such as Rubisco. An obvious explanation for the labeling of abundant proteins outside nucleotide binding sites would be that solvent-exposed lysine residues have a low basal reactivity toward AcATP and that this causes aspecific labeling of highly abundant proteins. Strangely enough, however, AcATP labeling of abundant proteins nevertheless requires divalent ions and can be competed with nucleotides that do not carry an acyl reactive group. These data support the exciting idea that ATP-competable AcATP labeling sites might identify novel nucleotide binding sites in proteins. Interestingly, a recent crystal structure of rice Rubisco with NADPH revealed that two of the labeled lysines are in close proximity to the phosphate in NADPH (supplemental Fig. S4D) (47). Thus, labeling outside known nucleotide binding sites is an interesting topic for further studies.

**Future Prospects**—AcATP labeling is a powerful way to monitor ATP binding activities in native proteomes. Besides the majority of the 1099 protein kinases of Arabidopsis, AcATP can potentially label another 482 ATP binding proteins, representing another 11 different protein families. Although AcATP probes also label outside known nucleotide binding sites, the preferential labeling of nucleotide binding proteins in ATP binding pockets demonstrates the high selectivity of the AcATP probe. An interesting future prospect will be to study the changes of ATP binding activities of proteins upon an



external stimulus. Our data suggest that labeling not only reflects the abundance of the ATP binding proteins, but also is affected by the ATP binding activity of the proteins. ATP binding can be dynamic and dependent on the cellular state. However, further studies on changes in ATP binding activities upon a stimulus must involve pair-wise comparisons of the same labeled peptide under different conditions, and this requires further development of quantitative methods such as stable isotope labeling of amino acids in cell culture and the use of isotope-labeled probes. We aim to use AcATP labeling in the near future to identify selective inhibitors of Arabidopsis protein kinases and to study the activation of protein kinases upon external stimuli.

**Acknowledgments**—The mass spectrometry proteomics data have been deposited to the Proteome XChange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [49] with the dataset identifier PXD000188.

\* This work was financially supported by the Max Planck Society and the Deutsche Forschungsgemeinschaft (DFG projects HO 3983/3-1 and 4-1), and COST action CM 1004. M.K. greatly acknowledges funding by an ERC Starting Grant (Grant No. 258413).

§ This article contains [supplemental material](#).

\*\* To whom correspondence should be addressed: Tel.: +49 221 5062 245, Fax: +49 221 5062 207, E-mail: [hoorn@mpipz.mpg.de](mailto:hoorn@mpipz.mpg.de).

## REFERENCES

1. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815
2. Lamesch, P., Berardini, T. Z., Li, D., Swarbrick, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., and Huala, E. (2012) The Arabidopsis Information Resource (TAIR): improve gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210
3. Shiu, S. H., and Bleecker, A. B. (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10763–10768
4. Rasmussen, M. W., Roux, M., Petersen, M., and Mundy, J. (2012) MAP kinase cascades in Arabidopsis innate immunity. *Frontiers Plant Sci.* **2**, 169
5. Cheng, S. H., Willmann, M. R., Chen, H. C., and Sheen, J. (2002) Calcium signaling through protein kinases. The Arabidopsis calcium-dependent protein kinase gene family. *Plant Physiol.* **129**, 469–485
6. Patricelli, M. P., Szardenings, K., Liyanage, M., Nomanbhoy, T. K., Wu, M., Weissig, H., Aban, A., Chun, D., Tanner, S., and Kozarich, J. W. (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350–358
7. Patricelli, M. P., Nomanbhoy, T. K., Wu, J., Brown, H., Zhou, D., Zhang, J., Jagannathan, S., Aban, A., Okerberg, E., Herring, C., Nordin, B., Weissig, H., Yang, Q., Lee, J. D., Gray, N. S., and Kozarich, J. W. (2011) In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem. Biol.* **18**, 699–710
8. Okerberg, E. S., Wu, J., Zhang, B., Samii, B., Blackford, K., Winn, D. T., Shreder, K. R., Burbaum, J. J., and Patricelli, M. P. (2005) High-resolution functional proteomics by active-site peptide profiling. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4996–5001
9. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sadarsanam, S. (2002) The protein kinase complement of the human genome. *Science* **298**, 1912–1934
10. Deng, X., Dzamko, N., Prescott, A., Davies, P., Liu, Q., Yang, Q., Lee, J. D., Patricelli, M. P., Nomanbhoy, T. K., Alessi, D. R., and Gray, N. S. (2011) Characterization of a selective inhibitor of the Parkinson's disease kinase LRRK2. *Nat. Chem. Biol.* **7**, 203–205
11. Yang, Q., Deng, X., Lu, B., Cameron, M., Fearn, C., Patricelli, M. P., Yates, J. R., 3rd, Gray, N. S., and Lee, J. D. (2010) Pharmacological inhibition of BMK1 suppresses tumor growth through promyelocytic leukemia protein. *Cancer Cell* **18**, 258–267
12. Zhang, T., Inesta-Vaquera, F., Niepel, M., Zhang, J., Ficarro, S. B., Machleidt, T., Xie, T., Marto, J. A., Kim, N., Sim, T., Laughlin, J. D., Park, H., LoGrasso, P. V., Patricelli, M., Nomanbhoy, T. K., Sorger, P. K., Alessi, D. R., and Gray, N. S. (2012) Discovery of potent and selective covalent inhibitors of JNK. *Chem. Biol.* **19**, 140–154
13. Liu, Q., Kirubakaran, S., Hur, W., Niepel, M., Westhove, K., Thoreen, C. C., Wang, J., Ni, J., Patricelli, M. P., Vogel, K., Riddle, S., Waller, D. L., Traynor, R., Sanda, T., Zhao, Z., Kang, S. A., Zhao, J., Look, A. T., Sorger, P. K., Sabatini, D. M., and Gray, N. S. (2012) Kinome-wide selectivity profiling of ATP-competitive mammalian target of rapamycin (mTOR) inhibitors and characterization of their binding kinetics. *J. Biol. Chem.* **287**, 9742–9752
14. Vetter, M. L., Rodgers, M. A., Patricelli, M. P., and Yang, P. L. (2013) Chemoproteomic profiling identifies changes in DNA-PK as markers of early dengue virus infection. *ACS Chem. Biol.* **7**, 2019–2026
15. Adam, G. C., Burbaum, J., Kozarich, J. W., Patricelli, M. P., and Cravatt, B. F. (2004) Mapping enzyme active sites in complex proteomes. *J. Am. Chem. Soc.* **126**, 1363–1368
16. Chaudhry, C., Horwich, A. L., Brunger, A. T., and Adams, P. D. (2004) Exploring the structural dynamics of the *E. coli* chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. *J. Mol. Biol.* **342**, 229–245
17. Shek, P. Y. I., Zhao, J., Ke, Y., Siu, K. W. M., and Hopkinson, A. C. (2006) Fragmentation of protonated arginine, lysine and their methylated derivatives: concomitant losses of carbon monoxide or carbon dioxide and an amine. *J. Phys. Chem.* **110**, 8282–8296
18. Swiderski, M. R., and Innes, R. W. (2001) The Arabidopsis *PBS1* resistance gene encodes a member of a novel protein kinase subfamily. *Plant J.* **26**, 101–112
19. Wingenter, K., Trentmann, O., Wünsch, I., Hörmiller, I. I., Heyer, A. G., Reinders, J., Schultz, A., Geiger, D., Hedrich, R., and Neuhaus, H. E. (2011) A member of the mitogen-activated protein 3-kinase family is involved in the regulation of plant vacuolar glucose uptake. *Plant J.* **68**, 890–900
20. Gao, L., and Xiang, C. B. (2008) The genetic locus At1g73660 encodes a putative MAPKKK and negatively regulates salt tolerance in Arabidopsis. *Plant Mol. Biol.* **67**, 125–134
21. Silva, N. F., and Goring, D. R. (2002) The proline-rich, extension-like receptor kinase-1 (PERK1) gene is rapidly induced by wounding. *Plant Mol. Biol.* **50**, 667–685
22. Hanks, S. K., and Hunter, T. (1996) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576–596
23. Xing, W., Zhou, Y., Liu, Q., Liu, J., Luo, X., Huang, Q., Chen, S., Zhu, L., Bi, R., Hao, Q., Wu, J. W., Zhou, J. M., and Chai, J. (2007) The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature* **449**, 243–247
24. Park, C. J., and Ronald, P. C. (2012) Cleavage and nuclear localization of the rice XA21 immune receptor. *Nat. Commun.* **3**, 920
25. Melville, D. B. (1954) Biotin sulfoxide. *J. Biol. Chem.* **208**, 495–501
26. Upadhyay, K., Khattak, I. K., and Mullah, B. (2005) Oxidation of biotin during oligonucleotide synthesis. *Nucleosides Nucleotides Nucleic Acids* **24**, 5–7
27. Li, G., Meng, X., Wang, R., Mao, G., Han, L., Liu, Y., and Zhang, S. (2012) Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in Arabidopsis. *PLoS Genet.* **8**, e1002767
28. Bethke, G., Pecher, P., Eschen-Lippold, L., Tsuda, K., Katagiri, F., Glazebrook, J., Scheel, D., and Lee, J. (2012) Activation of the Arabidopsis thaliana mitogen-activated protein kinase MPK11 by the flagellin-derived elicitor peptide, flg22. *Mol. Plant Microbe Interact.* **25**, 471–480
29. Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J. E., and Innes, R. W. (2003) Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science* **301**, 1230–1233
30. DeYoung, B. J., Qi, D., Kim, S. H., Burke, T. P., and Innes, R. W. (2012) Activation of a plant nucleotide binding-leucine rich repeat disease protein by a modified self protein. *Cell. Microbiol.* **14**, 1071–1084

31. Lachaud, C., Prigent, E., Thuleau, P., Grat, S., Da Silva, D., Briere, C., Mazars, C., and Cotelle, V. (2013) 14-3-3-regulated  $\text{Ca}^{2+}$ -dependent protein kinase CPK3 is required for sphingolipid-induced cell death in Arabidopsis. *Cell Death Diff.* **20**, 209–217
32. Mori, I. C., Murata, Y., Yang, Y., Munemasa, S., Wang, Y. F., Andreoli, S., Tiriach, H., Alonso, J. M., Garper, J. F., Ecker, J. R., Kwak, J. M., and Schroeder, J. I. (2006) CDPKs CPK6 and CPK3 function in ABA regulation of guard cell S-type anion- and  $\text{Ca}^{2+}$ -permeable channels and stomatal closure. *PLoS Biol.* **4**, e327
33. Mehler, N., Wurzinger, B., Stael, S., Howmann-Rodrigues, D., Csaszar, E., Pfister, B., Bayer, R., and Teige, M. (2010) The  $\text{Ca}^{2+}$ -dependent protein kinase CPK3 is required for MAPK-independent salt-stress acclimation in Arabidopsis. *Plant J.* **63**, 484–498
34. Anthony, R. G., Khan, S., Costa, J., Pais, M. S., and Bogre, L. (2006) The Arabidopsis protein kinase PTI1-2 is activated by convergent phosphatidic acid and oxidative stress signalling pathways downstream of PDK1 and OXI1. *J. Biol. Chem.* **281**, 37536–37546
35. Haffani, Y. Z., Silva-Gagliardi, N. F., Sewter, S. K., Aldea, M. G., Zhao, Z., Nakhamchik, A., Cameron, R. K., and Goring, D. R. (2006) Altered expression of PERK receptor kinases in Arabidopsis leads to changes in growth and floral organ formation. *Plant Signal Behav.* **1**, 251–260
36. Finn, R. D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., and Bateman, A. (2010) The PFAM protein families database. *Nucleic Acids Res.* **38**, D211–D222
37. Plaxton, W. C. (1996) The organization and regulation of plant glycolysis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 185–214
38. Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L., and Söll, D. (2000) The adaptor hypothesis revisited. *Trends Biochem. Sci.* **25**, 311–316
39. Pujol, C., Bailly, M., Kern, D., Marechal-Drouard, L., Becker, H., and Duchene, A. M. (2008) Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNA<sub>Gln</sub> synthesis in plants. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6481–6485
40. Park, S., Rancour, D. M., and Bednarek, Y. S. (2008) In planta analysis of the cell cycle-dependent localization of AtCDC48A and its critical roles in cell division, expansion and differentiation. *Plant Physiol.* **148**, 246–258
41. Oqura, T., and Wilkinson, A. J. (2001) AAA+ superfamily ATPases: common structure- diverse function. *Gene Cells* **6**, 575–597
42. Sun, J., Ke, J., Johnson, J. L., Nikolau, B. J., and Wurtele, E. S. (1997) Biochemical and molecular biological characterization of CAC2, the Arabidopsis thaliana gene coding for the biotin carboxylase subunit of the plastidic acetyl-coenzyme A carboxylase. *Plant Physiol.* **115**, 1371–1383
43. Galperin, M. Y., and Koonin, E. V. (1997) A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci.* **6**, 2639–2643
44. Yamanaka, K., Ogura, T., Niki, H., and Hiraga, S. (1992) Identification and characterization of the smbA gene, a suppressor of the mukB null mutant of Escherichia coli. *J. Bacteriol.* **174**, 7517–7526
45. Liljelund, P., and Lacroute, F. (1986) Genetic characterization and isolation of the *Saccharomyces cerevisiae* gene coding for uridine monophosphokinase. *Mol. Gen. Genet.* **205**, 74–81
46. Holmes-Davis, R., Tanaka, C. K., Vensel, W. H., Hurkman, W. J., and McCormick, S. (2005) Proteome mapping of mature pollen of Arabidopsis thaliana. *Proteomics* **5**, 4864–4884
47. Matsumura, H., Mizohata, E., Ishida, H., Kogami, A., Ueno, T., Makino, A., Inoue, T., Yokota, A., Mae, T., and Kai, Y. (2012) Crystal structure of rice Rubisco and implications for activation induced by positive effectors NADPH and 6-phosphogluconate. *J. Mol. Biol.* **422**, 75–86
48. Baerenfaller, K., Grossmann, J., Grobei, M. A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., and Baginsky, S. (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. *Science* **320**, 938–941
49. Vizaino, J. A., Cole, R. G., Csordas, A., Dianas, J. A., Fabregot, A., Foster, J. M., Griss, J., Alpi, E., Birm, M., Conell, J., O'Kelly, G., Schoenegger, A., Ouellero, D., Perez-Riverol, Y., Resinger, F., Rios, D., Wang, R., and Hermajakob, H. (2013) The Proteomics Identifications (PRIDE) Database and Associated Tools: Status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069